Representing Interval Sets in a Computer

Jan Plaza
Computer Science Department
SUNY Plattsburgh

October 16, 2010
How to teach set operations?

- Consider finite and cofinite subsets of the set of rational numbers.

- They form a Boolean Algebra, representable in a computer.

- Not rich enough!
Plan

1. Define another Boolean Algebra (BA);

2. Design representation of sets in a computer
   – a canonical form of elements of this BA;

3. Formulate efficient algorithms for set operations.
Expressions vs. abstract objects

- **Expression denotes object.**

- To talk about expressions we quote them:
  
  \(\frac{1}{2}\) and \(0.5\) are two different expressions which denote the same rational number.

- To talk about objects denoted by expressions we do not quote:
  
  \(\frac{1}{2}\) and \(0.5\) are rational numbers and they are equal.

- Consider variables ranging over expressions.
  
  If variables \(a, b\) stand for \(\frac{1}{3}\) and \(\frac{2}{3}\) then \(\left[a, b\right)\) stands for \(\left[\frac{1}{3}, \frac{2}{3}\right)\).

- Quotation in which variables are expanded as above, is called **quasi-quotation**.

- These ideas come from Willard Van Orman Quine, 1940.
Our conventions

Instead of:

\[ \frac{1}{2} \]

we write:

expression \( \frac{1}{2} \)

and this represents quasi-quotation.
Why is this important?

Students, mathematicians, even computers work with expressions, not with mathematical objects.

An object of type Integer in computer memory should be thought of as an expression.

Other objects in computer memory can be expressions denoting sets.
Rational interval expressions

$a, b$ – expressions denoting rational numbers, $\text{eval}(a) < \text{eval}(b)$.

<table>
<thead>
<tr>
<th>Rational interval expression</th>
<th>Left endpoint</th>
<th>Right endpoint</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(a, b)$</td>
<td>$a$</td>
<td>$b$</td>
</tr>
<tr>
<td>$(a, b]$</td>
<td>$a$</td>
<td>$b$</td>
</tr>
<tr>
<td>$[a, b)$</td>
<td>$a$</td>
<td>$b$</td>
</tr>
<tr>
<td>$[a, b]$</td>
<td>$a$</td>
<td>$b$</td>
</tr>
<tr>
<td>$[a, a]$</td>
<td>$a$</td>
<td>$a$</td>
</tr>
<tr>
<td>$(-\infty, b)$</td>
<td>$-\infty$</td>
<td>$b$</td>
</tr>
<tr>
<td>$(-\infty, b]$</td>
<td>$-\infty$</td>
<td>$b$</td>
</tr>
<tr>
<td>$(a, \infty)$</td>
<td>$a$</td>
<td>$\infty$</td>
</tr>
<tr>
<td>$[a, \infty)$</td>
<td>$a$</td>
<td>$\infty$</td>
</tr>
<tr>
<td>$(-\infty, \infty)$</td>
<td>$-\infty$</td>
<td>$\infty$</td>
</tr>
</tbody>
</table>

Rational interval expressions denote **rational intervals**, as expected.
Interval precedence

Let $X, Y$ be rational intervals.
Define $X < Y$ iff for any $x \in X$ and for any $y \in Y$ we have $x < y$.

This concerns sets, not expressions.
Recall that a **Boolean algebra of sets** is any family of sets which contains the greatest set and is closed under operations of union, intersection and set-difference.

\[\mathbb{Z}\] – integers,
\[\mathbb{Q}\] – rational numbers,
\[\mathbb{A}\] – real algebraic numbers,
\[\mathbb{R}\] – real numbers.

Define **extended rational interval Boolean algebra** or **ERIBA** as the smallest Boolean algebra of sets containing sets \(\mathbb{Z}, \mathbb{Q}, \mathbb{A}\) and all the rational intervals.
Richness of ERIBA

The following sets belong to ERIBA.

- Any finite set of rational numbers.
- The set of irrational numbers: $\mathbb{R} - \mathbb{Q}$.
- The set of real transcendental numbers: $\mathbb{R} - \mathbb{A}$.
- The intersection of any rational interval with $\mathbb{Q}$ or $\mathbb{A}$ or the set of irrational numbers or the set of transcendental numbers.
- Any range of consecutive integers:
  \[
  \{m..n\} = [m, n] \cap \mathbb{Z}, \\
  \{m..\infty\} = [m, \infty) \cap \mathbb{Z}, \\
  \{-\infty..n\} = (-\infty, n] \cap \mathbb{Z}, \\
  \{-\infty..\infty\} = \mathbb{Z}.
  \]
ZQAR-chunk expressions

\[ Z = Z \]

\[ Q = Q - Z \]

\[ A = A - Q \]

\[ R = R - A \]

ZQAR-chunk expressions denote \textbf{ZQAR-chunks}, as expected.
ZQAR-chunk sets expressions denote **ZQAR-chunk sets**, as expected.
Operations on ZQAR-chunk sets

Let \( X_1 = (\sim \mathbb{Z} \cup \sim \mathbb{A}) = \langle \sim \mathbb{Z}, \sim, \sim \mathbb{A}, \sim \rangle = \langle 1010 \rangle. \)

Let \( X_2 = (\sim \mathbb{Q} \cup \sim \mathbb{A}) = \langle \sim, \sim \mathbb{Q}, \sim \mathbb{A}, \sim \rangle = \langle 0110 \rangle. \)

Then \( X_1 \cap X_2 = (\sim \mathbb{Z} \cup \sim \mathbb{A}) \cap (\sim \mathbb{Q} \cup \sim \mathbb{A}) = \sim \mathbb{A} = \langle \sim, \sim, \sim \mathbb{A}, \sim \rangle = \langle 0010 \rangle. \)

Intersection – bitwise conjunction,
Union – bitwise disjunction,
Complement – bitwise negation.
Rational ZQAR-interval expressions

Concatenate rational interval expression and ZQAR-chunk set expression. It denotes the intersection of the sets denoted by these expressions.

\[ \emptyset = (0, 5) \langle - , - , - , - , - \rangle \]
\[ \{a\} = [a, a] \langle \sim \mathbb{Z}, \sim \mathbb{Q}, - , - \rangle \text{ where } a \text{ is a fraction.} \]
\[ \{m..n\} = [m, n] \langle \sim \mathbb{Z}, - , - , - \rangle \]
\[ \{m..\infty\} = [m, \infty) \langle \sim \mathbb{Z}, - , - , - \rangle \]
\[ (a, b] = (a, b] \langle \sim \mathbb{Z}, \sim \mathbb{Q}, \sim \mathbb{A}, \sim \mathbb{R} \rangle \]
\[ \overline{A} = (-\infty, \infty) \langle - , - , - , - , \sim \mathbb{R} \rangle \text{ – real transcendental numbers.} \]
Rational ZQAR-intervals

Rational ZQAR-interval expressions denote rational ZQAR-intervals, as expected.

The family of rational ZQAR-intervals

- is a subset of ERIBA;
- contains elements ∅ and \( \mathbb{R} \);
- is closed under intersections: \( IX \cap JY = (I \cap J)(X \cap Y) \);
- is not closed under unions;
- is not closed under complements.
Normal components

1. By an **open normal component** we understand any of the following ZQAR-interval expressions:
   \((a, b)X,\)
   \((a, \infty)X,\)
   \((\infty, b)X,\)
   \((\infty, \infty)X,\)

2. By a **singleton normal component** we understand any ZQAR-interval expression of the following form:
   \([a, a]X.\)

3. By a **normal component** we understand either an open normal component or a singleton normal component.

A singleton normal component may denote the empty set.
Pre-normal forms

A pre-normal form is a finite sequence (list) of normal components $c_1, \ldots, c_n$ such that for any $i, j$ s.t. $1 \leq i < j \leq n$ we have $\text{base}(c_i) < \text{base}(c_j)$, written $c_1 \cup \ldots \cup c_n$.

A pre-normal form is called explicit if the union of the bases (i.e. rational intervals) of its components is $\mathbb{R}$.

For instance,
$$(-\infty, 0)\langle-,-,-,\rangle \cup [0, 0]\langle\tilde{\mathbb{Z}}, \tilde{\mathbb{Q}}, -,-\rangle \cup$$
$$\cup (0, 1)\langle-\mathbb{Q}, -\rangle \cup [1, 1]\langle-,-,-,-\rangle \cup (1, \infty)\langle-,-,-,-,-\rangle$$

is an explicit pre-normal form denoting ZQAR-interval $[0, 1]\langle-\mathbb{Q}, \mathbb{A}, -\rangle$.

In fact, for every rational ZQAR-interval there exists a pre-normal form denoting that ZQAR-interval.
Base and endpoints

The base of a rational ZQAR-interval expression $IX$ is set denoted by $I$.

The base of a pre-normal form, $\text{base}(c_1 \cup ... \cup c_n) = \text{base}(c_1) \cup ... \cup \text{base}(c_n)$.

Endpoints of a pre-normal form $f$, $\text{endpoints}(f)$ — the set of all finite endpoints of components of $f$.

These operations take expressions and return sets.

For instance, for

$$[0, 0]_{\langle \mathbb{Z}, \mathbb{Q}, -, - \rangle} \cup (0, 1)_{\langle -, \mathbb{Q}, \mathbb{A}, - \rangle} \cup [1, 1]_{\langle -, -, -, - \rangle} \cup (1, \infty)_{\langle -, -, -, -, - \rangle}$$

endpoints returns $\{0, 1\}$ and base returns $[0, \infty)$. 
Split operation

Let $A$ – finite set of normalized fractions,
Let $f$ - pre-normal form
$\text{split}_A(f)$ returns a pre-normal form.

Example
Let $A = \{\frac{1}{2}, 1, 2\}$.
Let $f = (0, 1)\langle \tilde{\mathbb{Z}}, \tilde{\mathbb{Q}}, \tilde{A}, - \rangle \cup [1, 1]\langle \tilde{\mathbb{Z}}, -, -, - \rangle$.
Then
$\text{split}_A(f) =$
$(0, \frac{1}{2})\langle \tilde{\mathbb{Z}}, \tilde{\mathbb{Q}}, \tilde{A}, - \rangle \cup [\frac{1}{2}, \frac{1}{2}]\langle \tilde{\mathbb{Z}}, \tilde{\mathbb{Q}}, \tilde{A}, - \rangle \cup$
$\cup (\frac{1}{2}, 1)\langle \tilde{\mathbb{Z}}, \tilde{\mathbb{Q}}, \tilde{A}, - \rangle \cup [1, 1]\langle \tilde{\mathbb{Z}}, -, -, - \rangle$
Intersection algorithm

Input. Explicit pre-normal forms \( f_1, f_2 \).

Output. An explicit pre-normal form denoting the intersection of the sets denoted by \( f_1 \) and \( f_2 \).

Termination. Always terminates.

Algorithm.

Compute \( f'_1 := \text{split}_{\text{endpoints}}(f_2)(f_1) \).

Compute \( f'_2 := \text{split}_{\text{endpoints}}(f_1)(f_2) \).

# Notice that \( f'_1, f'_2 \) are explicit pre-normal forms,

# Notice that \( \text{endpoints}(f'_1) = \text{endpoints}(f'_2) \).

Return: \( e_1 \cup \ldots \cup e_n \)

where \( e_i \) is the simplified expression \( c_i \cap d_i \)

where \( f'_1 \) is \( c_1 \cup \ldots \cup c_n \) and \( f'_2 \) is \( d_1 \cup \ldots \cup d_n \).
Other algorithms

One can also formulate algorithms for

- union,
- complement,
- set membership,
- subset,
- proper subset,
- equality of sets.
The algorithms are correct and can be implemented to have $O(n)$ time complexity where $n$ is the total number of normal components in the input.

The family of sets denoted by explicit pre-normal forms is closed under intersections, unions and complements with respect to $\mathbb{R}$.

Every pre-normal form denotes an element of ERIBA and every element of ERIBA is denoted by a pre-normal form.

The set of even natural numbers is not a member of ERIBA. The set of squares of natural numbers is not a member of ERIBA. The set of powers of 2 with natural exponents is not a member of ERIBA. The set of prime numbers is not a member of ERIBA.
An immediate generalization

- **B**: the set of all rational numbers which have finite representation base 2, (B stands for Binary),

- **D**: the set of all rational numbers which have finite representation base 10, (D stands for Decimal),

- **T**: the set of numbers of the form \( a + b\sqrt{2} \) where \( a, b \in \mathbb{Q} \), (T stands for Two),

- **S**: the set of real numbers which are solutions of quadratic equations, i.e. real numbers of the form \( (a + b\sqrt{c})/d \) where \( a, b, c, d \in \mathbb{Z}. \) (S stands for Square root.)

\[
\mathbb{Z} \subset \mathbb{B} \subset \mathbb{D} \subset \mathbb{Q} \subset \mathbb{T} \subset \mathbb{S} \subset \mathbb{A} \subset \mathbb{R}.
\]

All the constructions extend from rational ZQAR-intervals to rational ZBDQTSAR-intervals.
How to use this in teaching?

1. The paper can be read in an advanced undergraduate seminar on symbolic mathematical computation and computer algebra systems for math and computer science students.

2. The paper can be a base for a project on implementation of a computer algebra system that computes with sets.

3. The computer algebra system can be extended to an automated tutoring system that teaches beginning undergraduate students set operations.